

Manon Colin

PySpark Data Engineer

✉ manon.colin@example.com ☎ +33 6 412 72 31 📍 Montpellier

Summary

PySpark Data Engineer spécialisé dans traitement des capteurs de robots. Je construis des produits de données fiables avec PySpark, Delta Lake, Structured Streaming, Kafka, MLflow et relie les choix techniques à des résultats mesurables.

Work Experience

- PySpark Data Engineer**
Prairie Robotique · Montpellier

Jan 2022 - Present

 - Conçu des pipelines PySpark et Delta Lake pour traitement des capteurs de robots chez Prairie Robotique, avec les équipes produit et plateforme sur 3 400 unités robotisées; la part des traitements terminés à l'heure est passée de 92 % à 99,6 %. (rôle actuel, point 1)
 - Refondu les modèles Structured Streaming et le partitionnement du projet Feature lake des robots en encadrant une équipe de 9 ingénieurs; le temps médian des requêtes est descendu de 24 à 4,1 secondes. (rôle actuel, point 2)
 - Instrumenté les contrôles de schéma, de fraîcheur et de lineage sur 38 tables de production pour traitement des capteurs de robots; avec 14 analystes, les incidents de données ont baissé de 63 %. (rôle actuel, point 3)
 - Automatisé les backfills, alertes et runbooks des jobs Kafka à Montpellier pour 12 marchés; le support d'astreinte a économisé 86 heures par trimestre. (rôle actuel, point 4)

- Data Platform Engineer**
Prairie Robotique Data Studio · Montpellier

Apr 2019 - Dec 2021

 - Conçu des pipelines PySpark et Delta Lake pour traitement des capteurs de robots chez Prairie Robotique, avec les équipes produit et plateforme sur 3 400 unités robotisées; la part des traitements terminés à l'heure est passée de 92 % à 99,6 %. (rôle précédent, point 1)
 - Refondu les modèles Structured Streaming et le partitionnement du projet Feature lake des robots en encadrant une équipe de 6 ingénieurs; le temps médian des requêtes est descendu de 24 à 4,1 secondes. (rôle précédent, point 2)
 - Instrumenté les contrôles de schéma, de fraîcheur et de lineage sur 38 tables de production pour traitement des capteurs de robots; avec 5 équipes applicatives, les incidents de données ont baissé de 63 %. (rôle précédent, point 3)
 - Automatisé les backfills, alertes et runbooks des jobs Kafka à Montpellier pour 4 unités métier; le support d'astreinte a économisé 86 heures par trimestre. (rôle précédent, point 4)

Projects

Feature lake des robots

Création d'un produit de données documenté pour traitement des capteurs de robots, avec transformations testées et livraison prête pour l'analyse.

PySpark

Delta Lake

Structured Streaming

Kafka

Skills

Technologies

- PySpark
- Delta Lake
- Structured Streaming
- Kafka
- MLflow

Livraison

- Modélisation des données
- Qualité des données
- CI/CD
- Observabilité
- Documentation

Collaboration

- Ateliers de cadrage
- Réponse aux incidents
- Mentorat
- Livraison agile

Education

Université de Montpellier

Sep 2015 - Jun 2019

Licence informatique at Data engineering et systèmes logiciels

Certifications

Databricks Associate Developer for Spark

Organisme de certification professionnelle

May 2024

Data Reliability Workshop

Data Engineering Guild

Aug 2023

Languages

Français • Langue maternelle

Anglais • Niveau professionnel